

# Algoritma *Naïve Bayes* Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai ( Studi Kasus Kelurahan Utama )

*Naive Bayes Algorithm For Classification of Non-Cash Food Aid Recipients (Case Study of Utama Village)*

Castaka Agus Sugianto<sup>1</sup>, Firdi Rizky Maulana<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika - Politeknik TEDC Bandung

E-mail: <sup>1</sup>castaka@poltektedc.ac.id, <sup>2</sup>firdirizki@rocketmail.com

## Abstrak

Kelurahan Utama merupakan instansi pemerintahan di Cimahi Selatan. Kelurahan Utama menjalankan program pemerintah yaitu program Bantuan Pangan Non Tunai, dalam menjalankan program Bantuan Pangan Non Tunai sebagian warga banyak yang mengeluh karena tidak mendapat bantuan, sedangkan ada beberapa warga yang dianggap mampu justru mendapatkan bantuan. Berdasarkan latar belakang tersebut maka Penelitian ini melakukan proses pengolahan data menggunakan *data mining* untuk mengklasifikasi penerima dan bukan penerima bantuan pangan non tunai dengan metode klasifikasi menggunakan Algoritma *Naïve Bayes* dan Algoritma *Decision Tree* sebagai pembanding. Diharapkan data yang dihasilkan dari proses *data mining* bisa menjadi bahan evaluasi untuk pemerintah. Dalam penelitian ini mengklasifikasi data penerima dan bukan penerima bantuan pangan non tunai menggunakan teknik klasifikasi pada *data mining* menggunakan Algoritma *Naïve Bayes* dan Algoritma *Decision Tree* sebagai pembanding. Model *data mining* di buat menggunakan *RapidMiner*, dengan hasil nilai Probabilitas untuk class "PENERIMA" yaitu 0,481 dengan pembulatan nilai menjadi 0,48 dan nilai Probabilitas untuk class "Bukan Penerima" yaitu 0,519 dengan pembulatan nilai menjadi 0,52. Algoritma *Naïve Bayes* mempunyai tingkat *Accuracy* sebesar 58,29%, *Precision* 92,90%, *Recall* 21,84%, *AUC* 0,765, *F-Measure* 34.42%. Sedangkan algoritma *Decision Tree* mempunyai tingkat *Accuracy* sebesar 73,97%, *Precision* 85,04%, *Recall* 61,92%, *AUC* 0,746, *F-Measure* 71,17%. Dalam hasil pengujian *T-Test* antara Algoritma *Naive Bayes* dan Algoritma *Decision Tree* didapat  $\alpha \leq 0.000$ , maka dapat disimpulkan pengujian *T-Test* antara Algoritma *Naïve Bayes* dan Algoritma *Decision Tree* hasilnya signifikan.

Kata kunci: Kelayakan, *Data Mining*, Klasifikasi, *T-Test*, *Naïve Bayes*, *Decision tree*

## Abstract

*Utama villlage is a government agency in southern cimahi. Utama village run the government program, namely the Non-Cash Food Aid program, in carrying out the Non-Cash Food Aid program, many residents complain that they do not get assistance, while there are some people who are considered able to get help. Based on this background, carry out data processing using data mining to classify recipients and not recipients of non-cash food aid with classification methods using the Naïve Bayes Algorithm and Decision Tree Algorithm as a comparison. It is hoped that the data generated from the data mining process can be an evaluation material for the government. In this study the authors classify recipient data and not recipients of non-cash food aid using classification techniques in data mining using the Naïve Bayes Algorithm and Decision Tree Algorithm as a comparison. The data mining model is made using RapidMiner, with the result that the Probability value for the class " RECEIVER " is 0.481 with rounding the value to 0.48 and the Probability value for the class 'Non Recipient' which is 0.519 with rounding the value to 0.52. The Naïve Bayes algorithm has an Accuracy level of 58.29%, Precision 92.90%, Recall 21.84%, AUC 0.765, F-Measure 34.42%. While the Decision Tree algorithm has an Accuracy level of 73.97%, Precision 85.04%, Recall 61.92%, AUC 0.746, F-Measure 71.17%. In the results of testing the T-Test between the Naive Bayes Algorithm and the Decision Tree Algorithm,  $\alpha \leq 0.000$  was obtained, it can be*

concluded that the T-Test between the Naïve Bayes Algorithm and the Decision Tree Algorithm was significant.

Keywords: Feasibility, Data mining, Classification, T-Test, Naïve Bayes, Decision Tree.

## 1. PENDAHULUAN

Masyarakat miskin adalah suatu kondisi dimana fisik masyarakat yang tidak memiliki akses ke prasarana dan sarana dasar lingkungan yang memadai, dengan kualitas perumahan dan pemukiman yang jauh di bawah standar kelayakan serta mata pencaharian yang tidak menentu yang mencakup seluruh multidimensi, yaitu dimensi politik, dimensi sosial, dimensi lingkungan, dimensi ekonomi dan dimensi aset[1].

Dalam membantu meminimalkan kemiskinan, kementerian sosial membuat program bantuan untuk warga miskin, bantuan yang dibuat kementerian sosial adalah Bantuan Pangan Non Tunai (BPNT). Bantuan Pangan Non Tunai (BPNT) adalah bantuan sosial pangan yang disalurkan dalam bentuk non tunai dari pemerintah kepada Keluarga Penerima Manfaat (KPM) setiap bulannya. Melalui mekanisme uang elektronik yang digunakan hanya untuk membeli bahan pangan di pedagang bahan pangan atau disebut *E-warong* yang bekerjasama dengan Bank Penyalur [2].

Kelurahan adalah pembagian wilayah administratif di Indonesia di bawah kecamatan. Kelurahan juga mempunyai tugas untuk memberikan pelayanan kepada masyarakat dalam hal pembuatan *E-KTP*, pembuatan Kartu Keluarga, Akta Kelahiran, dan pengajuan bantuan program dari pemerintah. Salah satu program dari pemerintah adalah program Bantuan Pangan Non Tunai yang di selenggarakan pada tahun 2017 di 44 Kota di Indonesia[1]. Salah satunya adalah Kota Cimahi, dan disebarkan ke beberapa kelurahan yang ada di kota Cimahi, salah satunya adalah Kelurahan Utama.

Kelurahan Utama adalah suatu instansi pemerintahan yang berada di dalam cakupan Kecamatan Cimahi Selatan, dalam menjalankan tugas sebagai pelayanan publik, Kelurahan Utama memberikan pengajuan dari beberapa Program Pemerintah, seperti Kartu Indonesia Sehat, Kartu Indonesia Pintar, BPJS dan Bantuan Pangan Non Tunai. Salah satu program yang memiliki kendala di Kelurahan Utama adalah Program Bantuan Pangan Non Tunai.

Dari hasil final data kependudukan 2017 jumlah warga di Kelurahan Utama sebanyak 36.054 KK yang terbagi dalam 16 RW dan 91 RT. Dari data tersebut tercatat jumlah warga miskin (gakin) sebanyak 1086 KK. Dari data warga miskin tersebut, penerima BPNT sebanyak 626 KK.

Dengan adanya masalah tersebut, mengakibatkan adanya ketidak sinkronan penerima BPNT. Masyarakat yang berhak menerima bisa jadi tidak menerima, sedangkan warga yang bukan merupakan warga miskin malah justru menerima BPNT. Jika penerima BPNT jelas terdata dengan baik, serta syarat kelayakan penerimaan BPNT yang terdata dengan baik, maka akan mengurangi ketidaksinkronan penerima BPNT dan juga lebih memudahkan petugas di dalam pembagian BPNT tersebut. Oleh karena itu dibutuhkan analisis terhadap penerima BPNT yang bisa dilakukan dengan berbagai teknik analisis, salah satunya menggunakan teknik *data mining*.

*Data mining* adalah suatu cara yang bertujuan dalam penemuan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam database atau sumber data lain yang dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai aturan proses [3]. Salah satu teknik *data mining* adalah fungsi klasifikasi. Fungsi klasifikasi dapat memudahkan mengidentifikasi kelayakan penerima dan bukan penerima BPNT. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target [4]. Beberapa Algoritma dalam klasifikasi diantaranya adalah *C4.5*, *ID3*, *K-Nearest Neighbor Classifier*, *Naive Bayes Classifier*, *SVM*, dan *ANN* [5].

*Naive bayes Classifier* merupakan salah satu metode yang digunakan dalam data mining yang didasarkan pada keputusan *bayes*. *Naive bayes Classifier* memiliki kemampuan klasifikasi seperti metode *decision tree* dan *neural network*. metode ini dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [6]. *Naive Bayes* merupakan algoritma klasifikasi yang ada dalam teknik data mining yang menggunakan teori bayes [7]. *Naive Bayes* memiliki tingkat akurasi yang lebih baik dibanding model *classifier* lainnya [8]. *Naive Bayes* adalah salah satu algoritma klasifikasi yang simple dan mudah dalam implementasinya [9]. Ketika diuji menggunakan data yang benar algoritma ini sangat efektif, apalagi ketika di kombinasikan dengan pemilihan fungsi, sehingga *redundant* pada

data dapat dikurangi [10], Algoritma *naïve bayes* memiliki *performance* yang bagus ketika digabungkan dengan metode *clustering* [11]. *Naïve Bayes* sudah terbukti memiliki akurasi yang tinggi dibandingkan dengan algoritma *support vector machine*[12]. Hal ini sejalan dengan penelitian terdahulu dalam memprediksi data *customer* dengan metode *naïve bayes* mendapatkan hasil yang baik [13]. Maka penelitian ini akan menggunakan *data mining* teknik klasifikasi dengan algoritma *Naïve Bayes*.

Selain itu, penelitian yang sudah dilakukan oleh Andi Bekto Rahardjo [14] Penerapan Data Mining untuk Mengklasifikasi Penerima dan Bukan Penerima Kartu Identitas Miskin (KIM), hasilnya dari data yang berjumlah 240 record dan 15 atribut menghasilkan tingkat akurasi tertinggi yaitu 93,06%. Penelitian lain mengenai klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kab. Gorontalo mendapat hasil akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai Precision sebesar 92% dan Recall sebesar 86%.[15] Sedangkan yang hamper sama namun beda algoritma yang digunakan yaitu K-nearest Neighbor untuk mengklasifikasikan penerima Kartu Indonesia menggunakan variabel yang sudah ditentukan mendapatkan hasil yang baik.[16]. Senada yang dilakukan oleh Yeni Magfiroh, algoritma *Naive Bayes* memiliki *performance* yang tinggi[17].

## 2. METODE PENELITIAN

Ada 9 (sembilan) tahapan yang akan dilakukan Penelitian ini. Diantaranya:

### 2.1 Menentukan Masalah

Menentukan Masalah yang merupakan proses awal dalam penelitian ini. Proses ini sudah dijelaskan pada latar belakang mengenai klasifikasi penerima dan bukan penerima bantuan pangan non tunai di kelurahan utama.

### 2.2 Menentukan Tujuan dan Ruang Lingkup

Menentukan tujuan dan ruang lingkup merupakan apa yang ingin di capai dalam penelitian, serta batasan yang dipakai agar penelitian tidak terlalu melebar.

### 2.3 Mencari Literature

Mencari *literature* merupakan mencari informasi yang berkaitan dengan penelitian seperti *data mining*, *algoritma Naïve Bayes*, *C4.5*, dan klasifikasi kelayakan.

#### 2.3.1 Pencarian Literature

Metode pengumpulan *literature* mempunyai peranan penting dalam penelitian ini, karena dari metode pengumpulan data akan menghasilkan informasi untuk prediksi yang akan dihasilkan dan di kumpulkan untuk mendukung selama proses penelitian.

#### 2.3.2 Study Pustaka

Metode yang di lakukan dengan mencari sumber dari buku-buku, jurnal, dan media internet yang berhubungan dengan penelitian ini. Data-data yang kumpulkan dari hasil studi pustaka adalah:

1. Materi tentang *data mining*.
2. Pengumpulan jurnal-jurnal yang berhubungan dengan *data mining*, Algoritma *Naïve Bayes*, Algoritma *C4.5*.
3. Pencarian indikator yang sesuai.
4. Pencarian teori-teori yang di butuhkan yang berkaitan dengan penelitian ini.

### 2.4 Menentukan Indikator Data

Indikator yang di gunakan dalam data pada penelitian menggunakan indikator yang di gunakan oleh Badan Pusat Statistik Kota Cimahi.

## 2.5 Pengumpulan Data

Proses pengumpulan data dilakukan dengan cara mengambil data dari daftar kolektif hasil Penerima Bantuan Pangan Non Tunai Kelurahan Utama dan data hasil kuisioner PPLS ( Program Perlindungan Sosial ) Tahun 2011.

## 2.6 Preparasi Data

Proses preparasi data yang di lakukan adalah dengan mengubah dari data *hard copy* kedalam format *excel*.

### 2.6.1 Data Yang Digunakan

Data yang di gunakan dalam penelitian ini di ambil dari hasil final data kependudukan 2017 jumlah warga di Kelurahan Utama sebanyak 36.054 KK yang terbagi dalam 16 RW dan 91 RT. Dari data tersebut tercatat jumlah warga miskin (gakin) sebanyak 1086 KK. Dari data warga miskin tersebut, penerima BPNT sebanyak 626 KK. Data tersebut di ambil dari awal bulan desember 2017 sampai bulan Mei 2018. Data dalam penelitian ini akan di uji dengan Algoritma *Naïve bayes* dan di bandingkan dengan dengan algoritma *C4.5* dan di validasi menggunakan *cross validation*, *T-test*, serta *Anova*. Berikut *dataset* daftar kolektif hasil penyebaran Kuisioner bisa di lihat pada tabel 1.

Gambar 1 Tabel Dataset Data Kuisioner PPLS Tahun 2011

NO	KODE_PROV	NAMA_PROV	KODE_KAB	NAMA_KAB	KODE_KEC	NAMA_KEC	KODE_DESA	NAMA_DESA	KODE_SLS	NAMA_SLS	NO_ID	JUD_DOK	NO_ID	JML_KLG	JML_ART	B3K1	B3K2	B3K3	B3K4	B3K5	B3K6	B3K7	B3K8	B3K9	B3K10	B3K11	B3K12A	B3K12B	B3K12C	B3K12D	B3K12E	B3K12F	B3K12G	B3K12H	B3K12I	B3K13A	B3K13B	B3K13C	B3K13D	B3K13E	B3K13F	KETERANGAN			
1.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	HERM	LS	1000	1	7	6	20	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA		
2.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	UTEP	LS	1000	1	5	5	56	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
3.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	FAHRL	LS	1000	1	8	5	56	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
4.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	MANA	LS	1000	1	4	5	56	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
5.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	AMIN	LS	1000	1	5	5	28	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
6.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	SARTIF	LS	1000	1	3	5	10	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
7.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	AHMA	LS	1000	1	6	6	56	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
8.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	AHMA	08	2000	1	2	6	99	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
9.32	JAWA	77	KOTA	010	CIMA	003	UTAN	002	RW	0 0521	ERWIN	LS	P327	1	4	6	28	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
10.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	SUPRI	LS	1000	1	5	5	18	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
11.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	RUSLI	LS	1000	2	6	5	49	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
12.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	CUMAL	LS	1000	1	7	5	28	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
13.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	ENTIS	LS	1000	1	5	6	42	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
14.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	SUGIH	LS	1000	1	4	4	9	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
15.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	SAEPUL	LS	1000	1	4	6	12	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
16.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	KARLI	LS	1000	1	1	6	9	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
17.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	WARIS	LS	1000	1	5	6	21	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
18.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	SUBAFI	LS	1000	1	4	5	30	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
19.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	MAMFI	LS	1000	1	3	1	70	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
20.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	ASEP	LS	1000	1	3	1	35	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
21.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	TOMI	LS	1000	1	4	1	24	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
22.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	TONGEL	LS	1000	1	5	6	42	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
23.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	NAND	LS	1000	1	5	1	36	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
24.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	UANC	LS	1000	1	4	6	9	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
25.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	DANIS	LS	1000	1	4	5	12	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
26.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	OTAHUS	LS	1000	1	6	4	28	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
27.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	TATI	LS	2000	1	3	6	15	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
28.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	YAYATI	LS	P327	1	4	5	36	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
29.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	WIHEL	LS	P327	1	4	5	49	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	PENERIMA
30.32	JAWA	77	KOTA	010	CIMA	003	UTAN	003	RW	0 0511	SUPRI	LS	P327	1	3	5	42	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA
31.32	JAWA	77	KOTA	010	CIMA	003	UTAN	007	RW	0 0511	SOPAN	LS	1000	1	1	6	4	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	BUKAN PENERIMA

### 2.6.2 Proses Cleansing Data

Sebelum dilakukan proses *data mining* pada proses ini *dataset* akan di reparasi yaitu menghilangkan *field* yang tidak perlu di pakai ketika proses *data mining*. Dalam penelitian ini padaawalnya terdapat 47 *field* kemudian setelah di lakukan *cleansing* data menjadi 31 *field* adapun *field*nya bisa di lihat pada Tabel 2.

Tabel 1 Tabel Keterangan Field

No	Atribut	Jenis Data	Type Data
1	NO	Integer	Id
2	NAMA_KRT	Varchar	Attribute
3	JML_KLG	Integer	Attribute
4	JML_ART	Integer	Attribute
5	B3K1	Integer	Attribute

6	B3K2	<i>Integer</i>	<i>Attribute</i>
7	B3K3	<i>Integer</i>	<i>Attribute</i>
8	B3K4A	<i>Integer</i>	<i>Attribute</i>
9	B3K4B	<i>Integer</i>	<i>Attribute</i>
10	B3K5A	<i>Integer</i>	<i>Attribute</i>
11	B3K5B	<i>Integer</i>	<i>Attribute</i>
12	B3K6	<i>Integer</i>	<i>Attribute</i>
13	B3K7	<i>Integer</i>	<i>Attribute</i>
14	B3K8A	<i>Integer</i>	<i>Attribute</i>
15	B3K8B	<i>Integer</i>	<i>Attribute</i>
16	B3K9	<i>Integer</i>	<i>Attribute</i>
17	B3K10	<i>Integer</i>	<i>Attribute</i>
18	B3K11	<i>Integer</i>	<i>Attribute</i>
19	B3K12A	<i>Integer</i>	<i>Attribute</i>
20	B3K12D	<i>Integer</i>	<i>Attribute</i>
21	B3K12E	<i>Integer</i>	<i>Attribute</i>
22	B3K12G	<i>Integer</i>	<i>Attribute</i>
23	B3K12H	<i>Integer</i>	<i>Attribute</i>
24	B3K12I	<i>Integer</i>	<i>Attribute</i>
25	B3K13A	<i>Integer</i>	<i>Attribute</i>
26	B3K13B	<i>Integer</i>	<i>Attribute</i>
27	B3K13C	<i>Integer</i>	<i>Attribute</i>
28	B3K13D	<i>Integer</i>	<i>Attribute</i>
29	B3K13E	<i>Integer</i>	<i>Attribute</i>
30	B3K13F	<i>Integer</i>	<i>Attribute</i>
31	KETERANGAN	<i>Varchar</i>	<i>Nominal</i>

Variabel yang di pakai dalam pengumpulan data menggunakan Variabel dari Badan Pusat Statistik, pemilahan dan pengkodean *Field* dapat dijelaskan dengan di keterangan tabel 3.

Tabel 2 Keterangan Variabel Data

<b>Blok Keterangan</b>	<b>Variabel Data</b>	<b>Nama Field</b>
1.	Status Penguasaan Bangunan Tempat Tinggal yang ditempati	B3K1
2.	Luas Lantai.... $m^2$	B3K2

3.	Jenis Lantai Terluas	B3K3
4.	a. Jenis Dinding Terluass b. Jika 4a Berkode 1 atau 2, Kondisi Dinding	B3K4A B3K4B
5.	a. Jenis Lantai Terluas. b. Jika 5a berkode 1,2,3,4 atau 5 Kondisi Atap	B3K5A B3K5B
6.	Sumber Air Minum	B3K6
7.	Cara Memperoleh Air Minum	B3K7
8.	a. Sumber Penerangan Utama b. Jika Listrik PLN (8a = 1), daya terpasang	B3K8A B3K8B
9.	Bahan bakar/energy utama untuk memasak	B3K9
10.	Penggunaan Fasilitas tempat buang Air Besar	B3K10
11.	Tempat Pembuangan Akhir Tinja	B3K11
12.	Apakah rumah tangga memiliki sendiri asset sebagai berikut : a. Mobil. d. Sepeda Motor. e. Sepeda. g. Lemari es/kulkas. h. Tabung gas 12kg atau lebih. i. HP	B3K12A B3K12D B3K12E B3K12G B3K12H B3K12I
13.	Apakah rumah tangga menjadi peserta program berikut : a. Program Keluarga Harapan (PKH) b. Beras Untuk Orang Miskin (Raskin) c. Jaminan Kesehatan Masyarakat (Jamkesmas) d. Asuransi Kesehatan Lainnya. e. Jaminan Sosial Tenaga Kerja (Jamsostek). f. Keluarga Berencana (KB)	B13KA B13KB B13KC B13KD B13KE B13KF

## 2.7 Proses dan Hasil

Proses data mining yang di lakukan menggunakan algoritma *Naïve Bayes* yang menghasilkan nilai akurasi, dan akan di bandingkan dengan algoritma *C4.5* karena algoritma *C4.5* merupakan algoritma klasifikasi yang dapat membuat pohon keputusan untuk menghasilkan data.

## 2.8 Analisis

Proses analisis pada penelitian ini mengamati dan menjelaskan hasil dari proses *RapidMiner*.

## 2.9 Penarikan Kesimpulan

Penarikan Kesimpulan merupakan hasil dari proses pengujian yang telah dilakukan serta hasil analisis Penelitian ini.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengujian Menggunakan Algoritma Naïve Bayes

Pengujian ini dilakukan untuk mengetahui nilai *Accuracy*, *Precision*, *Recall*, *AUC*, *F-Measure* yang keluar serta mengetahui seperti apa hasil *Simple Distributor* dan Grafik ROC yang akan di hasilkan pada proses *data mining*.

Berikut adalah hasil pengujian menggunakan Algoritma Naïve Bayes dan menggunakan validasi *Cross Validation*. Dari Tabel 4. menunjukan bahwa Algoritma Naïve Bayes mendapatkan hasil *Accuracy* sebesar 58,29%, *Precision* 92,90%, *Recall* 21,84%, *AUC* 0,765, *F-Measure* 34,42%.

Hasil Pengujian Algoritma Naïve Bayes/					
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>	<i>F-Measure</i>
<i>Algoritma Naïve Bayes</i>	58,29%	92,90%	21,84%	0,765	34.42%

Gambar 2 Hasil Pengujian Algoritma Naïve Bayes

Untuk membuktikan hasil klasifikasi penerima Bantuan Pangan Non Tunai, bisa dilakukan dengan perhitungan manual *naïve bayes classifier* dengan rumus (1) sebagai berikut :

$$P(C_i | X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Menghitung jumlah kelas dari klasifikasi yang terbentuk

C1 = Penerima

C2 = Bukan Penerima

Perhitungan :

$$P(X|C_1) = P(\text{penerima}) = 301/626 = 0,481$$

$$P(X|C_2) = P(\text{bukan penerima}) = 325/626 = 0,519$$

Perhitungan :

$$P(\text{hakmilik bangunan="peribadi"} | \text{penerima}) = 98/301 = 0,325$$

$$P(\text{hakmilik bangunan="peribadi"} | \text{bukan penerima}) = 300/325 = 0,923$$

$$P(\text{luas bangunan="0-50 m2"} | \text{penerima}) = 102/301 = 0,339$$

$$P(\text{luas bangunan="0-50 m2"} | \text{bukan penerima}) = 38/325 = 0,117$$

$$P(\text{jenis lantai bangunan=">50% tanah"} | \text{penerima}) = 108/301 = 0,359$$

$$P(\text{jenis lantai bangunan=">50% tanah"} | \text{bukan penerima}) = 0/325 = 0$$

$$P(\text{sumber air minum="air tidak dilindung"} | \text{penerima}) = 65/301 = 0,216$$

$$P(\text{sumber air minum="air tidak dilindung"} | \text{bukan penerima}) = 30/325 = 0,092$$

$$P(\text{sumber penerangan="450 watt"} | \text{penerima}) = 207/301 = 0,688$$

$$P(\text{sumber penerangan="450 watt"} | \text{bukan penerima}) = 50/325 = 0,154$$

$$P(\text{bahan bakar="kompor minyak"} | \text{penerima}) = 1/301 = 0,003$$

$$P(\text{bahan bakar="kompor minyak"} | \text{bukan penerima}) = 0/325 = 0$$

$$P(\text{lantai ketersediaan wc="jamban umum"} | \text{penerima}) = 78/301 = 0,24$$

$$P(\text{lantai ketersediaan wc} = \text{"jamban umum"} \mid \text{bukan penerima}) = 0/325 = 0$$

$$P(\text{Kepemilikan Aset} = \text{"Mobil"} \mid \text{penerima}) = 1/301 = 0,003$$

$$P(\text{Kepemilikan Aset} = \text{"Mobil"} \mid \text{bukan penerima}) = 171/325 = 0,526$$

$$P(\text{Kepemilikan Aset} = \text{"Sepeda Motor"} \mid \text{penerima}) = 268/301 = 0,890$$

$$P(\text{Kepemilikan Aset} = \text{"Sepeda Motor"} \mid \text{bukan penerima}) = 290/325 = 0,892$$

$$P(\text{Kepemilikan Aset} = \text{"Sepeda"} \mid \text{penerima}) = 68/301 = 0,226$$

$$P(\text{Kepemilikan Aset} = \text{"Sepeda"} \mid \text{bukan penerima}) = 17/325 = 0,052$$

$$P(\text{Kepemilikan Aset} = \text{"Lemari Es"} \mid \text{penerima}) = 60/301 = 0,199$$

$$P(\text{Kepemilikan Aset} = \text{"Lemari Es"} \mid \text{bukan penerima}) = 301/325 = 0,926$$

$$P(\text{Kepemilikan Aset} = \text{"Tabung Gas"} \geq 12\text{kg} \mid \text{penerima}) = 6/301 = 0,020$$

$$P(\text{Kepemilikan Aset} = \text{"Tabung Gas"} \geq 12\text{kg} \mid \text{bukan penerima}) = 45/325 = 0,138$$

$$P(\text{Kepemilikan Aset} = \text{"HP"} \mid \text{penerima}) = 301/301 = 1$$

$$P(\text{Kepemilikan Aset} = \text{"HP"} \mid \text{bukan penerima}) = 325/325 = 1$$

$$P(\text{kepesertaan program} = \text{"PKH"} \mid \text{penerima}) = 20/301 = 0,066$$

$$P(\text{kepesertaan program} = \text{"PKH"} \mid \text{bukan penerima}) = 171/325 = 0,526$$

$$P(\text{kepesertaan program} = \text{"Raskin"} \mid \text{penerima}) = 260/301 = 0,864$$

$$P(\text{kepesertaan program} = \text{"Raskin"} \mid \text{bukan penerima}) = 0/325 = 0$$

$$P(\text{kepesertaan program} = \text{"jamkesmas"} \mid \text{penerima}) = 89/301 = 0,296$$

$$P(\text{kepesertaan program} = \text{"jamkesmas"} \mid \text{bukan penerima}) = 312/325 = 0,96$$

$$P(\text{kepesertaan program} = \text{"Asuransi"} \mid \text{penerima}) = 5/301 = 0,017$$

$$P(\text{kepesertaan program} = \text{"Asuransi"} \mid \text{bukan penerima}) = 71/325 = 0,218$$

$$P(\text{kepesertaan program} = \text{"Jamsostek"} \mid \text{penerima}) = 102/301 = 0,339$$

$$P(\text{kepesertaan program} = \text{"Jamsostek"} \mid \text{bukan penerima}) = 210/325 = 0,646$$

$$P(\text{kepesertaan program} = \text{"KB"} \mid \text{penerima}) = 296/301 = 0,983$$

$$P(\text{kepesertaan program} = \text{"KB"} \mid \text{bukan penerima}) = 305/325 = 0,938$$

Mengkalikan semua hasil dari atribut

$$P(X \mid \text{penerima}) = 0,325 \times 0,339 \times 0,359 \times 0,216 \times 0,688 \times 0,003 \times 0,24 \times 0,003 \times 0,890 \times 0,226 \times 0,199 \times 0,020 \times 1 \times 0,066 \times 0,864 \times 0,296 \times 0,017 \times 0,339 \times 0,983 =$$

$$P(X \mid \text{bukan penerima}) = 0,923 \times 0,117 \times 0 \times 0,092 \times 0,154 \times 0 \times 0 \times 0,526 \times 0,892 \times 0,052 \times 0,926 \times 0,138 \times 1 \times 0,526 \times 0 \times 0,96 \times 0,218 \times 0,646 \times 0,938 = 0$$

$$P(X \mid \text{penerima}) \times P(\text{penerima}) = 0 \times 0,481 = 0,000000000000000972$$

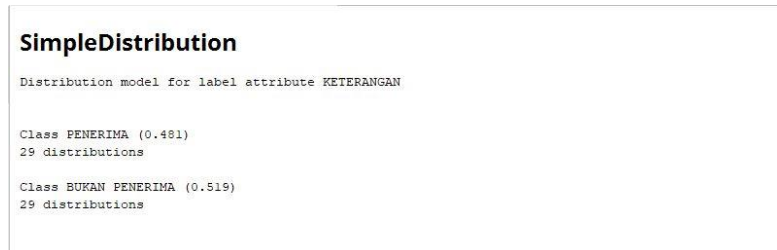
$$P(X \mid \text{bukan penerima}) \times P(\text{bukan penerima}) = 0 \times 0,519 = 0$$

Dari perhitungan diatas disimpulkan dengan data warga penerima Bantuan Pangan Non Tunai seperti diatas maka warga tersebut digolongkan dalam klasifikasi penerima Bantuan Pangan Non Tunai karena hasil probabilitas  $P(X \mid \text{penerima}) \times P(\text{penerima})$  lebih besar dari yang bukan penerima.



### 3.2 Simple Distribution Naïve Bayes

Hasil model distribusi Algoritma *Naïve Bayes* dari proses *data mining* digunakan untuk melihat nilai probabilitas dari Algoritma *Naïve Bayes*. Bisa dilihat nilai yang dihasilkan dari Algoritma *Naïve Bayes* nilai *class* “PENERIMA” sebanyak 0,481 sedangkan *class* “BUKAN PENERIMA” sebanyak 0,519, ditunjukkan pada gambar 1.



Gambar 3 Hasil Simple Distribution Algoritma *Naïve Bayes*

### 3.3 Pengujian Menggunakan Algoritma Decision Tree

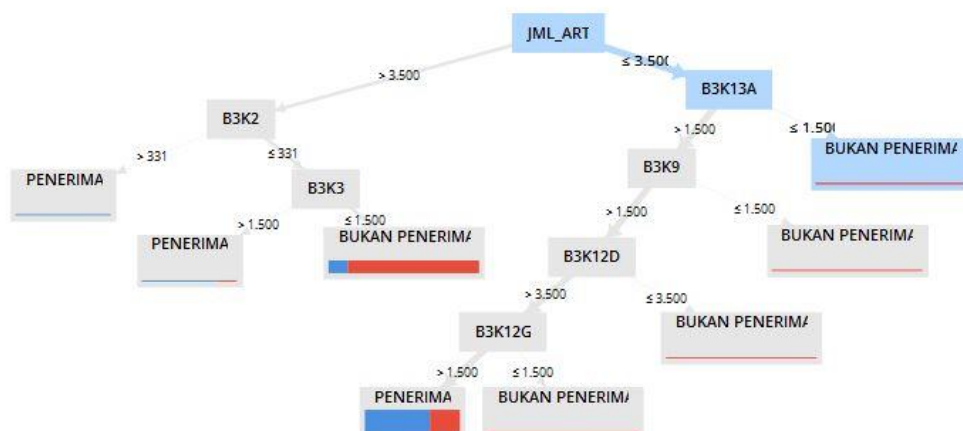
Pengujian ini dilakukan untuk mengetahui nilai *Accuracy*, *Precision*, *Recall*, *AUC*, *F-Measure* yang keluar serta mengetahui seperti apa hasil Pohon Keputusan yang akan di hasilkan pada proses *data mining*. Berikut adalah hasil pengujian menggunakan Algoritma *Decision Tree* dan menggunakan validasi *Cross Validation*. Dari Tabel 5 menunjukan bahwa Algoritma *Decision Tree* mendapatkan hasil *Accuracy* sebesar 73,97%, *Precision* 85,04%, *Recall* 61,92%, *AUC* 0,746, *F-Measure* 71,17%.

Hasil Pengujian Algoritma <i>Decision Tree</i>					
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>AUC</i>	<i>F-Measure</i>
<i>Algoritma Decision Tree</i>	73,97%	85,04%	61,92%	0,746	71,17%

Gambar 4 Hasil Pengujian Algoritma *Decision Tree*

### 3.4 Hasil Pohon Keputusan

Hasil pohon keputusan dari proses *data mining* digunakan untuk melihat apa saja indikator yang berpengaruh dalam data yang telah digunakan dalam penelitian ini, hasil pohon keputusan bisa di lihat pada gambar 2.



Gambar 5 Hasil Pohon Keputusan

### 3.5 Perbandingan Algoritma Naïve Bayes dan Algoritma Decision Tree Menggunakan T-Test

T-Test adalah metode pengujian hipotesis yang menggunakan objek penelitian dengan menggunakan dua perlakuan yang berbeda. Atau disebut juga uji beda untuk mengetahui adakah perbedaan pengujian antara 2 kelompok/pengujian menggunakan metode yang berbeda[18].

**H<sub>0</sub>**=tidak ada perbedaan hasil pengujian antara algoritma *decision tree* dan algoritma *naïve bayes*

**H<sub>a</sub>**= ada perbedaan hasil pengujian antara algoritma *decision tree* dan algoritma *naïve bayes*

Sebelum kita mentafsirkan atas hasil output maka mengetahui dasar pengambilan keputusan dalam uji t-test adapun kriteria pengujian :

**H<sub>0</sub>** diterima dan **H<sub>a</sub>** ditolak jika nilai  $\text{sig}/\alpha > 0,05$

**H<sub>0</sub>** ditolak dan **H<sub>a</sub>** diterima jika nilai  $\text{sig}/\alpha < 0,05$

A	B	C
	0.586 +/- 0.049	0.740 +/- 0.055
0.586 +/- 0.049		0.000
0.740 +/- 0.055		

Gambar 6 Hasil Pengujian T-Test

Berdasarkan hasil pengujian t-test yang ada pada tabel 6. bisa di lihat, hasil pengujian penelitian ini didapat  $\text{sig}/\alpha=0,000$ , jadi dalam penelitian ini hasilnya dapat di simpulkan signifikan/ada perbedaan pada pengujian *T-Test* antara Algoritma *Naïve Bayes* dan Algoritma *Decision Tree*.

## 4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan oleh Penelitian ini maka Penelitian ini menarik kesimpulan sebagai berikut :

1. Model *data mining* di buat menggunakan *RapidMiner*, dengan hasil nilai Probabilitas untuk class "PENERIMA" yaitu 0,481 dengan pembulatan nilai menjadi 0,48 dan nilai Probabilitas untuk class "BUKAN PENERIMA" yaitu 0,519 dengan pembulatan nilai menjadi 0,52.
2. Berdasarkan keseluruhan pengujian dan validasi Algoritma *Naïve Bayes* mendapatkan *Accuracy* sebesar 58,29%, *Precision* 92,90%, *Recall* 21,84%, *AUC* 0,765 dan *F-Measure* 34,42%, sedangkan Algoritma *Decision Tree* mendapatkan *Accuracy* sebesar 73,97%, *Precision* 85,04%, *Recall* 61,92%, *AUC* 0,746, *F-Measure* 71,17%. Dari hasil tersebut dapat dikatakan bahwa Algoritma *Decision Tree* lebih besar tingkat *Accuracy* dalam memprediksi data penerima dan bukan penerima bantuan pangan non tunai di kelurahan utama.
3. Berdasarkan hasil pengujian *T-Test* antara Algoritma *Naïve Bayes* dan Algoritma *Decision Tree* didapat  $\text{sig}/\alpha$  sebesar 0,000, maka dapat disimpulkan hasil pengujian *T-Test* antara Algoritma *Naïve Bayes* dan Algoritma *Decision Tree* hasil nya ada perbedaan yang signifikan.

## 5. SARAN

Beberapa saran dari Penelitian ini untuk penelitian selanjutnya adalah sebagai berikut :

1. Disarankan bagi penelitian selanjutnya sebaiknya membuat aplikasi untuk klasifikasi kelayakan penerima bantuan pangan non tunai di berbagai wilayah di Indonesia.
2. Untuk mendapatkan hasil yang lebih baik lagi kedepannya perlu dilakukan pengujian dengan data yang sama atau dengan Algoritma yang berbeda, beberapa Algoritma klasifikasi seperti *K-Nearest Neighbor*, *Neural Network*, *Support Vector Machine*, *ID3*, *CART*, *Linear Discriminant Analysis*.

## DAFTAR PUSTAKA

- [1] S. Nazara and S. K. Rahayu, "Program Keluarga Harapan (PKH): Program Bantuan Dana Tunai Bersyarat di Indonesia," 2013.
- [2] Kementrian Sosial, "Pedoman Umum Bantuan Pangan Non Tunai," 2017.
- [3] I. H. Witten, E. Frank, and M. A. Hall, *Data mining*. 2011.
- [4] C. Orsenigo and C. Vercellis, "Multicategory classification via discrete support vector machines," *Comput. Manag. Sci.*, vol. 6, no. 1, pp. 101–114, 2009.
- [5] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," *Orient. J. Comput. Sci. Technol.*, vol. 8, no. 1, pp. 13–19, 2015.
- [6] Larose, "An Introduction to Data Mining," *Trad. Adapt. Thierry Vallaud*, vol. 28, pp. 9–35, v, 2005.
- [7] B. Santoso, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [8] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009.
- [9] Fitriyani and R. S. Wahono, "Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes," *J. Softw. Eng.*, vol. 1, no. 2, pp. 101–108, 2015.
- [10] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. Elsevier, 2011.
- [11] C. A. Sugianto, "Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan," *Techno.com*, vol. 14, no. 4, pp. 336–342, 2015.
- [12] T. H. Apandi and C. A. Sugianto, "Analisis Komparasi Machine Learning Pada Data Spam Sms," *J. TEDC*, vol. 12, no. 1, pp. 58–62, 2018.
- [13] S. De Cnudde and D. Martens, "Loyal to your city? A data mining analysis of a public service loyalty program," *Decis. Support Syst.*, vol. 73, pp. 74–84, 2015.
- [14] A. B. Rahardjo, "Penerapan Data Mining Untuk Mengklasifikasi Penerima dan Bukan Penerima Kartu Identitas Miskin ( KIM ) Kelurahan Sumurrejo Gunungpati dengan Metode Naive Bayes Classifier," 2015.
- [15] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, p. 160, 2018.
- [16] A. Z. Abidin and Y. I. Kurniawan, "Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma K-Nearest Neighbor," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 4, no. 2, pp. 151–160, 2018.
- [17] Y. Magfiroh, "Rekomendasi Bantuan Penerima Bantuan Iuran (PBI) Menggunakan Klasifikasi Algoritma Naive Bayes," Universitas Nusantara PGRI Kediri, 2016.
- [18] Khafiizh Hastuti, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif," in *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012)*, 2012, vol. 14, no. 1, pp. 241–249.